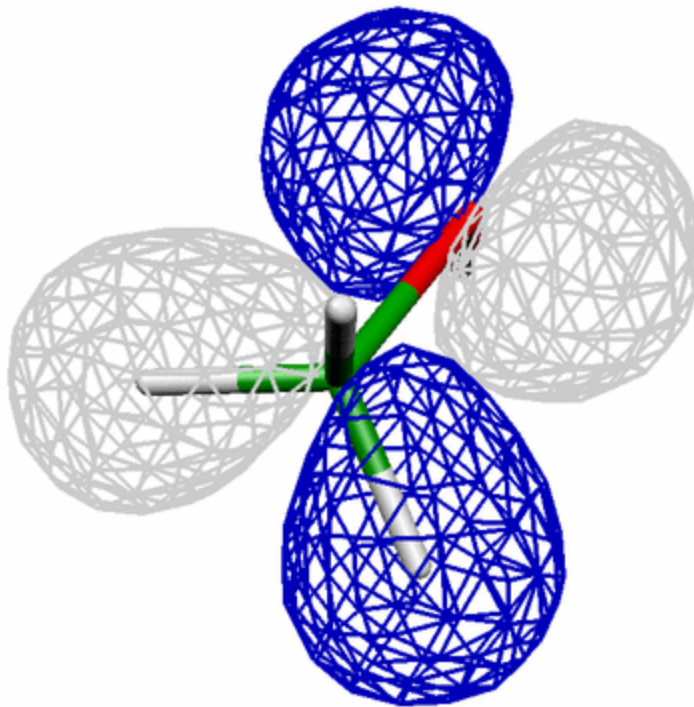


[Next](#) [Up](#) [Previous](#)Next: [Introduction](#)

University of Illinois at Urbana-Champaign  
Beckman Institute for Advanced Science and Technology  
Theoretical and Computational Biophysics Group

# Introduction to QM simulations



**Markus Dittrich**

May 29, 2003.

- 
- [Introduction](#)
  - [Part 1: Minimization and Structural Study of a Model System](#)
  - [Part 2: Determination of Proton Affinities](#)

---

*markus@ks.uiuc.edu*

[Next](#) [Up](#) [Previous](#)

**Next:** [Minimization and Structural Study of a Model System](#) **Up:** [Introduction to QM simulations](#)  
**Previous:** [Introduction to QM simulations](#)

## Subsections

- [Computational Setup and Preparations](#)
- [Setting up runs with GAMESS](#)
- [Understanding Z-matrices](#)

---

# Introduction

This hands-on session will introduce you to quantum mechanical (QM) simulation methods and provide you with a few examples illustrating the powerful tools QM simulations offer for biomolecular modeling. The first part constitutes a basic introduction to QM simulations: How to use the QM software, set up runs etc.. All the files that you need to have will be provided for you. You will be able to calculate some useful properties and then put them to work. Since QM calculations are often quite time consuming we will provide you with intermediate results, so that you can focus on the essentials and continue on, when you are stuck.

In the second part of the tutorial you will then use QM simulations to calculate the proton affinity of water and methanol, which constitutes a very important quantity of biomolecular systems.

Before submitting your first simulation you need to make yourself familiar with the quantum chemistry package that you will use to perform your computations. This will be the topic of the following section, where you will learn how to set up a QM calculation. You also need to copy some files into you working directory before you can get started.

## Computational Setup and Preparations

The quantum chemistry program that you will use to perform your calculations is called [GAMESS](#) and can be accessed via a shell-script called **runqms**. In order to make life easier for you we will provide a complete directory structure named QM\_tutorial containing all the files you will need. First go into the working directory where you will perform all the exercises for this tutorial. Hence type

```
tbss> cd ~/tbss.work
```

Next, copy the directory which contains the QM tutorial files into your working directory by typing

```
tbss> cp -rp TOP_DIR/sumschool03/tutorials/05-qm-tutorial/QM_tutorial ./
```

Here you have to replace `TOP_DIR` by the directory where the main `sumschool03` directory is located, e.g. `~/Desktop` if it is located on your desktop or `/mnt/cdrom/` if you copy it from CD.

Now you will have a directory `QM_tutorial` containing three subdirectories: one for part 1 and one for part 2 of the tutorial. The third one contains the GAMESS manual which will be helpful in case you would like to run your own simulations (you can view pdf files by using the command `acroread`

filename.pdf).

In the directory `part1` you will find sub-directories for each of the calculations you will perform in part 1. Please make sure to run each simulation in its appropriate directory in order to keep track of all the different results (Don't worry, we will show you how!). Each directory also has a sub-directory called `sample` which contains sample input and pre-calculated output files. This will allow you to check your results and to make sure that you have set up your simulation properly.

## Setting up runs with GAMESS

Just like NAMD2, the quantum chemistry package GAMESS is driven and controlled by an input file which tells the program about the system under investigation and what to do with it. This will be explained by means of the following example file, which, in fact, is the input file for your first real QM calculation using GAMESS. You don't need to worry about where the file is located until part 1 of the tutorial.

```
! Minimization of retinal analogue (6-31G)
$CONTRL SCFTYP=RHF RUNTYP=OPTIMIZE COORD=ZMT NZVAR=36
ICHARG=1 $END
$SYSTEM MEMORY=90000000 $END
$BASIS NGAUSS=6 GBASIS=N31 $END
$STATPT OPTTOL=1.0E-4 NSTEP=30 $END
$SCF DIRSCF=.TRUE. $END
$ELPOT IEPOT=1 WHERE=PDC $END
$PDC PTSEL=CONNOLLY CONSTR=CHARGE $END
$DATA
Minimization of retinal analogue
C1
H
N      1      1.008000
H      2      1.008000      1      120.000
C      2      1.300000      3      120.000      1      180.000
C      4      1.450000      3      109.471      2      180.000
C      5      1.335000      4      120.000      3      180.000
C      6      1.450000      5      109.471      4      180.000
C      7      1.335000      6      120.000      5      180.000
H      4      1.089000      3      109.471      2      300.000
H      5      1.089000      4      120.000      9      180.000
H      6      1.089000      5      109.471      10     120.000
H      7      1.089000      6      120.000      11     180.000
H      8      1.089000      7      109.471      12     120.000
H      8      1.089000      7      109.471      12     240.000
$END
```

We will now guide you through the file to help you understand what is being done. The GAMESS manual contains much more detail as well as additional options and you should have a look at it. The input file basically needs to specify at least three essential items: the molecular structure of the system under consideration, the quantum mechanical method ("level of theory") to solve the corresponding Schrodinger equation and the type of run, e.g. energy calculation, structure optimization etc..

The input file which controls the computation performed by GAMESS consists of a list of keyword groups. Each group starts with a \$ sign followed by its name (note the required single space in front of each of them), e.g. `$CONTRL`, and ends with `$END`. The following will explain the keyword groups relevant to you. For more information please refer to the GAMESS manual (you can open it by via `acroread manual/gamess.pdf&`).

**\$CONTRL:** As its name implies, this field controls the GAMESS run. The **SCFTYP** keyword tells GAMESS which wave function type to use and for you it will always be **SCFTYP=RHF** (RHF=Restricted Hartree-Fock). **RUNTYTYP** tells GAMESS what to do: single point energy calculation or, in your first example, optimization of the provided geometry via **RUNTYTYP=OPTIMIZE**. The next two keywords control how the coordinates of the molecule are input: **COORD=ZMT** tells GAMESS that you provide a so called Z-matrix (to be explained below) and **NZVAR=33** lists the number of internal coordinates (there are  $3N-6$ , where  $N$  is the number of atoms). **NZVAR=33** also tells GAMESS that you request the optimization to be performed in the internal coordinate space (as opposed to simple cartesian coordinates which you can request via **NZVAR=0**). The final item in the main control group, **ICHARG=1** provides GAMESS with the total charge of the system, 1 in this case.

**\$\$SYSTEM:** This group tells GAMESS the amount of core memory available to it. You don't need to worry about it!

**\$\$BASIS:** In this section you tell GAMESS what particular basis to use for your simulation. Initially you will be using the 6-31G basis, which is specified by the keyword sequence **NGAUSS=6** (i.e. 6-31G), **GBASIS=N31** (i.e. 6-31G).

**\$\$STATPT:** This group controls the properties of the geometry optimization run. With **OPTTOL** you specify the value of the gradient at which the minimization is considered to be converged. **NSTEP** is the maximum number of optimization steps allowed; it has to be large enough to accommodate the number of steps required for optimization of your system.

**\$\$SCF:** This option controls the specific way GAMESS performs its calculations and you don't need to worry about the details. Consult the documentation in case you're interested.

**\$\$ELPOT:** This keyword section requests the calculation of the electrostatic potential at points controlled by the input group **\$\$PDC**.

**\$\$PDC:** This group determines the points at which to calculate the electrostatic potential for the purpose of fitting atomic charges to this potential. These are the so called electrostatic potential **ESP** derived charges which you will need in the first part of the hands-on session.

**\$\$DATA:** This section provides GAMESS with the conformation of the system under consideration. In principle, there are several ways to provide this information: one could simply supply the list of atoms and their respective x,y and z coordinates, similar to a pdb file. Sometimes, however, it is more convenient to specify the molecular arrangement in terms of internal coordinates via a Z-matrix. In this case, rather than giving the xyz-coordinates of all atoms, one specifies the distances, angles and dihedrals between the system's atoms. The next section will provide you with a short tutorial on how to set up a Z-matrix.

A successful GAMESS run will create at least a \*.dat file (certain RUNTYTYPs create more files) in addition to the standard output which is typically stored in the \*.log file. The latter contains all the information that you need during this tutorial. The \*.dat file stores additional information, e.g. the wavefunction and other restart values important for more advanced simulations.

## Understanding Z-matrices

The purpose of this section is to help you understand how to find internal coordinates (i.e. distances, angles and dihedrals) that specify the conformation of the system under consideration and also to tell

you how to use them to set up a Z-matrix. It is very important to keep in mind that in contrast to simple xyz-coordinates there is generally no unique set of internal coordinates for a system. This implies that you have the possibility of choosing a set of internal coordinates that is well suited for your system. Particularly for structure optimization and transition state searches a "good" choice of internal coordinates can be crucial for successful and efficient completion of your run. This is related to the fact that some of the internal coordinates may be linearly dependent, making it impossible to vary them independently, thereby prohibiting the optimization procedure to reach the equilibrium structure efficiently. In this sense the internal coordinates provided in the sample input file are "bad" since they constitute a rather linearly dependent set. They are nevertheless good enough and illustrate well the concept of a Z-matrix. (If you have time during one of your evening sessions we encourage you to develop a better one and repeat the minimization of part 1 using e.g. cartesian (NZVAR=0) and your new internal coordinates. You will observe that the convergence of the optimization runs will greatly depend on the choice of coordinates.) Let us now look more closely at the **\$DATA** field of the input file:

```

$DATA
Minimization of retinal analogue
C1
H
N      1      1.008000
H      2      1.008000      1      120.000
C      2      1.300000      3      120.000      1      180.000
C      4      1.450000      3      109.471      2      180.000
C      5      1.335000      4      120.000      3      180.000
C      6      1.450000      5      109.471      4      180.000
C      7      1.335000      6      120.000      5      180.000
H      4      1.089000      3      109.471      2      300.000
H      5      1.089000      4      120.000      9      180.000
H      6      1.089000      5      109.471      10     120.000
H      7      1.089000      6      120.000      11     180.000
H      8      1.089000      7      109.471      12     120.000
H      8      1.089000      7      109.471      12     240.000
$END

```

The geometry input starts in line 3 of the **\$DATA** field (the first two lines, however, need to be present: the first one is a text comment, the second line provides the symmetry of the system). Each line specifies the location of an atom in terms of its position with respect to its molecular neighbors. The overall orientation of the molecule obviously does not matter. The first atom in the Z-matrix (a hydrogen atom) constitutes the origin of the internal coordinate system and hence no distances or angles to any other atoms need to be specified. The second item in the list is a nitrogen atom (N) whose position is completely determined by providing the distance from the chosen origin, the hydrogen of line 1. Hence

```

N      1      1.008000

```

tells GAMESS that N has a distance of 1.008 Å from atom number 1. Atom number 3 (H) can be uniquely defined by providing, e.g. the distance to atom 2 and the angle it forms with atoms 1 and 2. Therefore

```

H      2      1.008000      1      120.000

```

means that atom 3 (H) has a distance of 1.008 Å from atom 2 (N) and the angle between atoms 1-2-3 (H-N-H) is 120 degrees. All remaining atoms can then be specified in terms of a distance, an angle and a dihedral as is the case in lines 4 to 13, e.g.

```

C      2      1.300000      3      120.000      1      180.000

```

implies that atom 4 (C) is located 1.30 Å from atom 2 (N) and that the angle and dihedral between atoms 3-2-4 and 1-3-2-4 is 120 and 180 degrees respectively. Again, note that this assignment is not unique and one should always try to chose the representation which is most convenient for the problem at hand.

---

[Next](#) [Up](#) [Previous](#)

**Next: Up: Previous:**  
*markus@ks.uiuc.edu*

[Next](#) [Up](#) [Previous](#)**Next:** [Determination of Proton Affinities](#) **Up:** [Introduction to QM simulations](#) **Previous:** [Introduction](#)

## Subsections

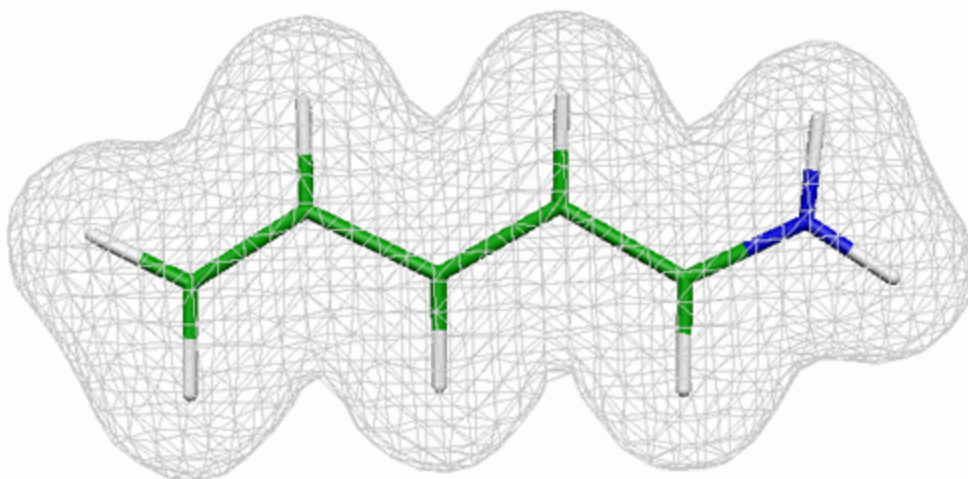
- [Introduction to the model system](#)
- [Submitting your first QM simulation](#)
- [Learning how to use the visualization tool MOLDEN](#)
- [Checking your run](#)
- [Looking at your trajectory](#)
- [Calculation of partial charges and force constants](#)
- [Putting your parameters to work](#)
- [Further studies](#)

---

# Part I: Minimization and Structural Study of a Model System

## Introduction to the model system

The model system that we will use to introduce you to quantum mechanical (QM) simulations is shown in the figure on the right (the gray wire



mesh surface indicates the electron density). The molecule is a retinal analogue and consists of one nitrogen, five carbon and eight hydrogen atoms, which are connected to each other in a conjugated polyene chain. The actual retinal molecule has a longer chain and is covalently bound to a lysine residue of its host protein via a Schiff base linkage. [Retinal](#) is the light absorbing moiety in the protein rhodopsin, which is responsible for color vision in the retina of our eyes.

## Submitting your first QM simulation

Before you can start studying the properties of the retinal analogue system you need to have a well minimized conformation of your system. You might e.g. be interested in the charge distribution and the equilibrium bond lengths. Since the optimization calculation can take up to 10 minutes you will first start the minimization run and use the meantime to acquaint yourself with the system. All the files you



need for part 1 are in the directory `QM_tutorial/part1/`. Go there now

```
tbss> cd part1
```

All paths in the remainder of part 1 will be relative to this directory! To start your minimization `cd` into the directory `min631G`.

```
tbss> cd min631G
```

Your working directory should contain the file `retmin.inp` (and sample output files in the `sample` subdirectory). The Z-matrix in the `$DATA` field of `retmin.inp` contains the starting conformation of the retinal analogue which was created "by hand" using standard values for bonds, angles and dihedrals. It does, therefore, not represent the conformation of minimal energy. This is the input file we looked at in the introduction section (have a look with `nedit` or your favorite editor!). Before you start the minimizations, open a second terminal window and go into your working directory. You can now start the minimization using the command

```
tbss> runrms retmin.inp >& retmin.log
```

The simulation will start running and your terminal will become inactive until the computation is finished. The simulation should be finished in less than 15 minutes. Make sure to check, in case the shell is still unresponsive after this amount of time has passed. In your second terminal window you should `cd` into the same directory where you started the GAMESS run and use the command

```
tbss> tail -f retmin.log
```

to view the last 10 lines of the log file (use `CNTRL C` to exit the tail command). If your job runs correctly you will see a large amount of output which is being generated.

NOTE: If you are running GAMESS via

```
tbss> runrms filename.inp >& filename.log
```

in a directory that already contains the files `filename.log`, `filename.dat` or `filename.F*` (e.g. from a previous run) it will terminate. You have to either remove or rename all files besides `filename.inp` to submit a new simulation in this particular directory. You can remove files with the `rm filename` command.

Now go back into your main directory by typing

```
tbss> cd ..
```

## Learning how to use the visualisation tool MOLDEN

Since the minimization will take a few minutes let us have a look at the retinal analogue molecule. You will not use [VMD](#) for this purpose, but rather the quantum chemistry visualization tool [MOLDEN](#) to look at GAMESS output.

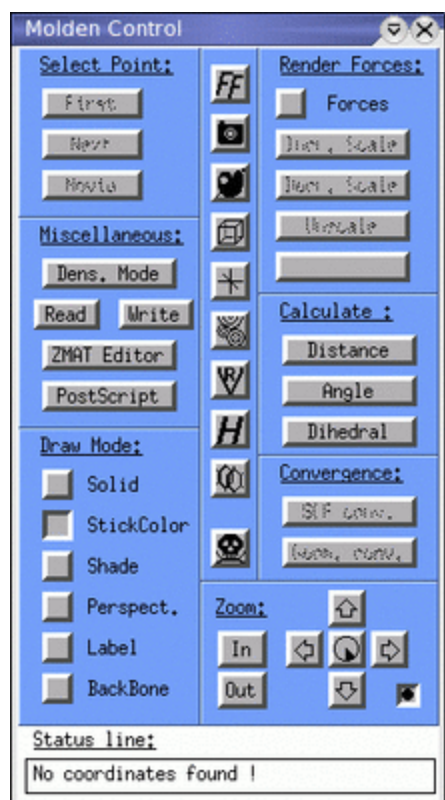
First, go into the directory `QM_tutorial/part1/pdbfile`, which contains the `pdb` file of the structure you are currently minimizing. Type

```
tbss> cd pdbfile
```

Then start up MOLDEN by typing

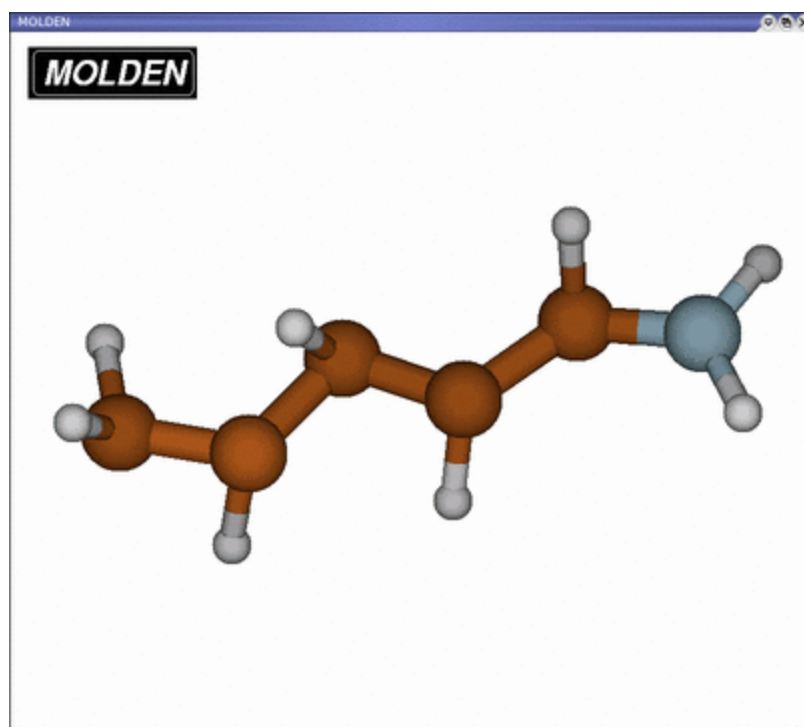
```
tbss> molder
```

Note that MOLDEN will "hijack" the terminal in which it was started, until you terminate your MOLDEN session. After typing the command you should see two windows: The main MOLDEN window and the control panel shown below.



To load the pdb file of the retinal analogue into MOLDEN press the **READ** button. This will open a file dialogue box where you can select the file `ret.pdb`. Please click on it. The molecule should show up in your main window and you can now close the dialogue box. The nice thing about MOLDEN is that it is also able to read the log file output generated by GAMESS and display geometries, trajectories and other useful information. The buttons in the **Draw Mode** section allow you to adjust the view to the way you prefer it. Try the **Solid**, **StickColor** and **Shade** buttons to obtain a view you like. Using your mouse in the MOLDEN window you can rotate the molecule. If you press the lower right button in the **Zoom** menu you are in the "sticky mode" and the molecule rotation will now follow your mouse. Take a few minutes to get acquainted to MOLDEN so you know what you can do with it. Try to avoid pressing the icons in the central column unless you know what you would like to do.

Once you feel comfortable with MOLDEN have a closer look at your system. After you loaded the pdb file of the retinal analogue into MOLDEN your main window will show something like below (yours may look different depending on your **Draw Style** and the way you rotated the molecule!)



The retinal analogue consists of 5 carbon, 1 nitrogen and 8 hydrogen atoms. To practice your understanding of internal coordinates you can use MOLGEN's geometry tools to confirm the Z-matrix input of the **\$DATA** field as discussed in a previous section. Hence measure some of the bonds, angles and dihedrals and compare with the values given in the input file. Note that the starting conformation of the retinal analogue is non-planar.

Now quit MOLGEN by hitting the SKULL button in the lower center of the control panel. Then go back into the main directory by typing

```
tbss> cd ..
```

## Checking your run

Once your simulation has finished, the shell in which you started the job will become responsive again. Before you do anything else you have to make sure that your simulation terminated normally and everything went smooth (it is a very common mistake to skip this step and take an improperly converged result as final result; always check your log file thoroughly!). First check your log file for successful job completion (you can use any editor of your choice; however, you need to know how to use it to search for strings, since you have to go through the rather lengthy log files. In nedit you can search with the Search -> Find menu item. Xemacs also has a simple interface to search a file in case you don't know how this feature works in your editor: You can either go to Edit -> Search or press CTRL S and then enter the string you are looking for at the bottom of the window). The last couple of lines should look similar to

```
EXECUTION OF GAMESS TERMINATED NORMALLY HH:MM:SS DD-MMM-19YY

DATA SERVER STATS: TOTAL DISTRIBUTED MEMORY USED (MEMDDI)=      0 MWORDS.
FIRST DATA SERVER'S MAXIMUM MEMORY=                0 WORDS, CPU=    0.0
SECONDS.
```

```
ddikick: all processes have ended gracefully.
unset echo
----- accounting info -----
Wed May 21 17:43:37 CDT 2003
Files used on the master node dakar.ks.uiuc.edu were:
ESC[00m-rw-r--r--    1 markus  tbres      257429 May 21 17:43 ESC[00m.//rethess.
-rw-r--r--    1 markus  tbres        1522 May 21 15:39 ESC[00m.//rethess.F05ESC[
-rw-r--r--    1 markus  tbres     1456752 May 21 17:43 ESC[00m.//rethess.F10ESC[
-rw-r--r--    1 markus  tbres     3100768 May 21 17:41 ESC[00m.//rethess.F22ESC[
-rw-r--r--    1 markus  tbres        1522 May 21 15:39 ESC[00m.//rethess.inpESC[
-rw-r--r--    1 markus  tbres       29985 May 21 17:43 ESC[00m.//rethess.ircESC[
-rw-r--r--    1 markus  tbres     562043 May 21 17:43 ESC[00m.//rethess.logESC[
ESC[m7101.060u 29.011s 2:04:02.62 95.8% 0+0k 0+0io 5617pf+0w
```

Most important is the statement `EXECUTION OF GAMESS TERMINATED NORMALLY` telling you that everything went fine (at least from the computational point of view, this does not imply that your simulation actually accomplished what you intended it to do). If your simulations did not terminate with this message you have to figure out what went wrong (contact your lab administrator if you are not sure what to do). Since you performed a geometry optimization you also have to check if GAMESS was able to find the equilibrium geometry or if, for some reason, this was not the case. In the log file look for the line

```
1      ***** EQUILIBRIUM GEOMETRY LOCATED *****
```

telling you that the geometry optimization was successful. If your simulation was not able to converge to the equilibrium geometry, try to figure out why. Maybe the maximum number of optimization steps requested via `NSTEP` in the input file was not sufficiently large. After you convinced yourself that your simulation terminated properly your working directory should contain the files `retmin.inp`, `ret_analogue.log` and `retmin.dat`

## Looking at your trajectory

We can now use MOLDEN to view the output generated by GAMESS during minimization. Make sure that you are in the directory `min631G` where you performed your minimization. Start up MOLDEN (click [here](#) if you are not sure how) and load the log file. MOLDEN allows you to view the sequence of geometries computed during minimization until the minimum was reached; use the three buttons in the **Select Point** menu: **FIRST** brings you to the first frame, **NEXT** to next one and **MOVIE** plays the whole trajectory. Look at your output and observe the changes in structure as the system approaches its minimal energy conformation. What is the major change and how could you explain this behavior. Try to identify which bonds in the system are double and which are single bonds. Any pattern?

## Calculation of partial charges and force constants

Now that you have obtained the optimized structure of the retinal analogue you can use GAMESS to determine some of the molecular properties you might be interested in. Let us assume that you would like to perform an MD simulation of rhodopsin, which contains retinal in its binding pocket. As you know by now, in order to do so you need to have the force field parameters of retinal, e.g. partial charges, bond lengths, force constants etc.. These are not provided in the standard CHARMM parameter files and you therefore may have to calculate them yourself. The optimized structure of the retinal analogue already provides the equilibrium values for bond lengths, angles and dihedrals as well as values for charges as you will see shortly, but you are still missing all the force constants. During this

session you will only consider how to obtain the force constants for the bonds and the partial charges of the atoms. The other parameters like the angle bending terms can also be calculated, but are more laborious. (You may want to use one of your evening labs to work on them. How would you e.g. go about calculating the stiffness of the angles? You may also want to explore how inclusion of electron correlation will affect the results.)

In order to obtain the force constants for the bonds you will have to calculate the Hessian matrix of the retinal analogue in its minimized conformation. This matrix consists of the second derivatives of the potential energy with respect to the coordinates. Its eigenvectors are the normal modes of the molecule. The normal modes can be projected onto the bonds of the molecule to extract their respective vibration frequencies  $f$  and force constants  $k$ , which are related via  $k = 4\pi^2 f^2 \mu$ , where  $\mu$  is the reduced mass (i.e.  $\mu = [m_1 m_2] / [m_1 + m_2]$ ) of the bonded atoms.

Hence, you will have to set up the run for the calculation of the hessian matrix. Please go into the directory `hess631G` to set up this run and perform your calculations. To obtain the input file for the calculation you can start from the input file you used for the minimization by copying it to a new file `rethess.inp`.

```
tbss> cd ../hess631G
tbss> cp ../min631G/retmin.inp ./rethess.inp
```

Open the new file `rethess.inp` in a text editor and delete the old internal coordinates in the **\$DATA** section (make sure to only delete the coordinates, not the first two non-coordinate lines in the file!). Next you have to insert the new, optimized internal coordinates. They can be extracted from the log file of your minimization run (remember: this file is located in the directory `min631G`). Hence load `retmin.log` into your favorite editor and search for the line

```
1      ***** EQUILIBRIUM GEOMETRY LOCATED *****
```

in the last third of the output file. Then scroll down a little more until you see the line

```
THE CURRENT FULLY SUBSTITUTED Z-MATRIX IS.
```

The next 13 lines represent your optimized Z-matrix. Now copy and paste the lines into your new input file `rethess.inp`. Finally you have to tell GAMESS that you would like to calculate the Hessian matrix and that you request the projection of the force constants onto the bonds. The first can be accomplished by changing the `runtyp` to **RUNTYP=HESSIAN** in the **\$CONTRL** group. The projection can be requested by inserting the keyword fields

```
$FORCE PURIFY=.TRUE. PRTIFC=.TRUE. DECOMP=.TRUE. $END
$ZMAT IZMAT(1)=1,1,2, 1,2,3, 1,2,4, 1,4,5, 1,5,6, 1,6,7,
                1,7,8, 1,4,9, 1,5,10, 1,6,11, 1,7,12, 1,8,13,
                1,8,14,
                2,1,2,3, 2,3,2,4, 2,3,4,5, 2,4,5,6, 2,5,6,7,
                2,6,7,8, 2,3,4,9, 2,4,5,10, 2,5,6,11,
                2,6,7,12, 2,7,8,13, 2,7,8,14,
                3,1,3,2,4, 3,2,3,4,5, 3,3,4,5,6, 3,4,5,6,7,
                3,5,6,7,8, 3,2,3,4,9, 3,9,4,5,10,
                3,10,5,6,11, 3,11,6,7,12, 3,12,7,8,13
                3,12,7,8,14
$END
```

at the end of the input file. Don't forget about inserting a single space in front of the \$. Once you are done you may compare your file with the templates in the `sample` directory. Make sure that you have no additional empty lines, tabs etc., otherwise GAMESS might get confused. Now you are ready to submit your simulation to calculate the Hessian matrix of the optimized retinal analogue. In the directory `hess631G` type

```
tbss> rungms rethess.inp >& rethess.log
```

While the simulation is running you should start extracting the atomic charges from the log file of your minimization run. What you need to know are the partial atomic charges of each of the atoms in the system. Quantum mechanics, however, only provides the electronic wavefunction and has no concept of assigning electrons to particular atoms. There is, therefore, no unique way of assigning partial charges with the help of QM calculations and several schemes exist to achieve this task. We will compare the results of three of them, namely Mulliken, Lowdin and ESP charges. The latter are obtained by fitting charges to the electrostatic potential, whereas the first two follow from projecting the electron density onto the nuclei. To extract the Mulliken and Lowdin charges from the log file of your minimization run, open `retmin.log` in a text editor. Since these charges are calculated at each step of the minimization run you have to make sure to pick the ones for the final minimization step. Hence start to search from the end of the log file and look for

```
TOTAL MULLIKEN AND LOWDIN ATOMIC POPULATIONS
```

Right underneath this line you will find two columns containing the Mulliken and Lowdin charges. If you are not sure that you found the correct entry you can peek [here](#). Make sure to pick the charges and not the populations. The ESP charges you will find almost at the end of the log file below the line titled

```
ELECTROSTATIC POTENTIAL
```

(check [here](#) if you are not sure). Now you have three sets of partial atomic charges for your system. You should compare them and notice that they are similar, but far from being identical.

Partial charges for force fields are typically determined using so called restrained ESP charges at a HF/6-31G\* level of theory, which are closely related to the ESP charges calculated by us. The reason being the fact that they are much less sensitive to a change in basis than the Mulliken or Lowdin charges. You can convince yourself that this is the case by using one of the evening sessions (don't do it now) to repeat the charge calculation using the HF/6-31+G level of theory, which includes diffuse functions (consult the manual to find the correct keyword. HINT: look in the **BASIS** section). You may also compare your computed set of partial charges with CHARMM values which were used in an actual MD simulation ([CHARMM charges for retinal analogue](#)).

Once the calculation of your Hessian matrix has finished have a look at the results. Make sure, as in the case of the minimization, that your GAMESS run has executed normally and nothing unexpected has happened. (note that this time you do not need to check for `1 ***** EQUILIBRIUM GEOMETRY LOCATED *****` since you did not perform an optimization, but a calculation of the Hessian matrix!) Use MOLDEN to look at the output file `rethess.log`. Hence start up MOLDEN, load the file and adjust the view of the molecule to your liking. Then press the **Norm Mode** button and a window will pop up that will allow you to view the normal modes of the retinal analogue that you have just calculated. The modes are ordered according to increasing frequencies (i.e. energy) and you can view them by simply clicking on them.

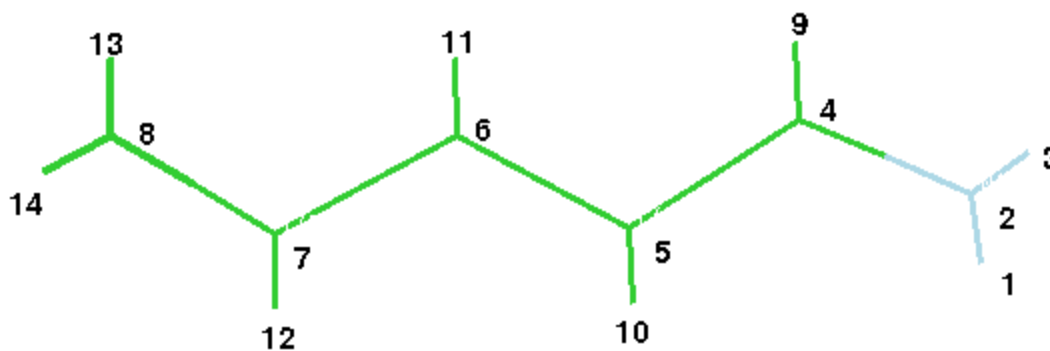
You should try to answer the following questions: Why do the first 6 normal modes have zero frequency and what are they? Can you discern qualitatively what distinguishes the low frequency modes (i.e. low

energy) from the high frequency ones (i.e. high energy)?

In order to extract the force constants for the bonds which follow from the hessian matrix you just calculated, have a look at the log file `rethess.log`. Hence, load the file into an editor and look for the lines

```
-----
INTERNAL COORDINATE                INTRINSIC FORCE CONSTANTS
                                   (HARTREES/BOHR**2)          (MDYN/ANG)
-----
```

The table right underneath lists the force constants for the 13 bonds in the retinal analogue in the same order as in the Z-matrix in the **\$DATA** section of the input file (note that e.g. `STR. 1 2` corresponds to the bond between atom 1 and 2). The figure below gives the atom numbering



They are given in  $\text{Hartree}/\text{Bohr}^2$  and you should convert them into  $\text{kcal}/(\text{mol}\text{\AA})^2$  to be able to use them in a CHARMM parameter file (note: 1 Hartree = 627.53 kcal/mol, 1 Bohr = 0.529\AA). Your calculation of bond frequencies/force constants is based on an harmonic approximation for the potential (remember, the Hessian is the second derivative of the potential energy function!). This generally overestimates the vibration frequencies and workers in the field often correct this by multiplying the frequencies  $f$  with by factor of 0.9 (what do you need to multiply the force constants with, considering that  $k = 4\pi^2 f^2 \mu$  ?)

Lastly, and quite unfortunately, the CHARMM force constants are defined via  $V_{\text{bond}} = k(x - x_0)^2$ , therefore lacking the conventional factor of one half. Hence, you also have to multiply your force constants by a factor of 0.5 before you can use them in a CHARMM parameter file. You can either do the calculations by hand using a calculator, or copy and paste the 13 lines with the force constants into a text editor and save them to the file `force.dat`. You can then use the command

```
tbss> gawk '{print $4*0.5*0.81*627.53/(0.529^2)}' force.dat > force_final.dat
```

to obtain the file `force_final.dat` with the converted force constants. Use the above figure of the retinal analogue to assign the force constants to the respective bonds and best write them into the figure. Next use MOLDEN to measure the bond distances and also enter them into the figure. Finally, put down the ESP charges of each atom. You have now collected all the information you need to complete the topology and parameter files for the retinal analogue. There is one last thing that you need to consider.

In the actual parameter file the bonds between atoms 8-13 and 8-14 have to be the same due to symmetry (the same is true for the bonds between 1-2 and 2-3!). There will hence only be one entry for each type of bond (this is why in the parameter file there are only eleven bond entries to complete and not 13). Since you probably have different values for the force constants and lengths of the bonds 8-13 and 8-14 just take the average! Finally, go back into the main directory

```
tbss> cd ..
```

## Putting your parameters to work

Finally, after you put all that effort into calculating the minimum energy conformation, ESP partial atomic charges, bond lengths, angles etc. you can now put them to use and complete the CHARMM parameter and topology files for our retinal analogue. After generating the psf file using psfgen you can then use NAMD to classically minimize the original conformation of the retinal analogue using your newly designed force field.

Hence go into the directory `namdmin/toppar`

```
tbss> cd namdmin/toppar
```

which contains the parameter and topology files for the retinal analogue (the values you need to fill in are denoted with Xs). Then fill in your calculated ESP charges (they go into the topology file as you know by now; note that the indices of the atoms in the topology file differs slightly from the ones in the Z-matrix and the above figure should help you to identify which correspond to which), the bond lengths and finally the force constants for the bonds that you have calculated (they go into the parameter file! If you have doubts about the correctness of your newly designed files you may peek at the provided files in the `sample` directory. If your values differ slightly, that's all right!) You don't have to worry about the angle and dihedral terms: leave them as they are. Once you have accomplished this, go into the directory `namdmin/psf` directory

```
tbss> cd ../psf
```

and create the psf file of the retinal analogue (remember: you need the original pdb file and the topology file that you just created; hence copy them into the directory).

```
tbss> cp ../toppar/top_ret.inp ./
tbss> cp ../../pdbfile/ret.pdb ./
```

Open VMD and use psfgen to generate the psf file (you should be able to this yourself by now; in case you run into serious trouble generating the psf file you may use the `create.png` script in the `sample` directory). Remember: To do so you need the completed topology file as well as the original pdb file. Please do not close VMD since we will need it again in a minute.

You are now ready to employ NAMD to minimize the retinal analogue structure using your own parameter set (1000 minimization steps are more than sufficient). Please do this in the directory `namdmin/min`, hence

```
tbss> cd ../min/
```

We encourage you to set up the NAMD input file yourself, but you may also use the provided one in the `sample` directory. You also need to copy the pdb and psf files from the `create_psf` directory as well as



the completed parameter file from the directory `toupper` into this directory. Then start the simulation by typing:

```
tbss> namd2 [name of inputfile] >& [name of outputfile]
```

After your minimization has finished, have a look at your final structure and compare with the QM optimization. How did your parameters perform?

## Further studies

The following provides some suggestions for further study during your evening sessions. Please don't do the exercises during the tutorial, but rather go on with part 2. We encourage you to repeat part 1 using a different, preferably larger, basis set, e.g. HF/6-31G\* or HF/6-31+G\*. In order to speed up your computations you can start your minimization from the optimized 6-31G geometry. Examine the changes in energy, structure and parameters (charges, bond lengths, force constants) when using a different basis set. This will show you the importance of carefully choosing the basis set of QM calculations in order to obtain meaningful results. It is also important to consider the influence of electron correlation by repeating the above calculations using density functional theory (DFT). Since this will be much more time consuming you will have to do this during one of the evening sessions. In order to employ DFT you simply insert the following line into you input file

```
$DFT DFTTYP=B3LYP $END
```

which will use the B3LYP functional to take into account electronic correlation effects.

---

[Next](#) [Up](#) [Previous](#)

**Next: Up: Previous:**  
*markus@ks.uiuc.edu*

TOTAL MULLIKEN AND LOWDIN ATOMIC POPULATIONS				
ATOM	MULL. POP.	CHARGE	LOW. POP.	CHARGE
1 H	0.560593	0.439407	0.709925	0.290075
2 N	7.851297	-0.851297	7.294316	-0.294316
3 H	0.559569	0.440431	0.713843	0.286157
4 C	5.705316	0.294684	5.802118	0.197882
5 C	6.284903	-0.284903	6.231322	-0.231322
6 C	6.060163	-0.060163	5.920342	0.079658
7 C	6.200227	-0.200227	6.167269	-0.167269
8 C	6.280680	-0.280680	6.083619	-0.083619
9 H	0.708499	0.291501	0.832090	0.167910
10 H	0.743739	0.256261	0.845828	0.154172
11 H	0.751188	0.248812	0.851069	0.148931
12 H	0.755467	0.244533	0.851334	0.148666
13 H	0.759243	0.240757	0.840730	0.159270
14 H	0.779115	0.220885	0.856195	0.143805

-----  
ELECTROSTATIC POTENTIAL  
-----

MERZ-KOLLMAN RADII USED FOR CHARGE FITTING  
NUMBER OF POINTS SELECTED FOR FITTING = 727

## NET CHARGES:

ATOM	CHARGE	E.S.D.
H	0.4611	0.0036
N	-0.8434	0.0114
H	0.4703	0.0035
C	0.5423	0.0125
C	-0.5439	0.0124
C	0.2023	0.0124
C	-0.2285	0.0120
C	-0.1814	0.0101
H	0.1403	0.0038
H	0.2662	0.0037
H	0.1403	0.0038
H	0.1902	0.0037
H	0.1984	0.0031
H	0.1860	0.0032

-----

ATOM NUMBER		CHARMM charge
1	H	0.505
2	N	-0.727
3	H	0.505
4	C	0.362
5	C	-0.180
6	C	0.131
7	C	-0.110
8	C	-0.090
9	H	0.181
10	H	0.180
11	H	0.110
12	H	0.100
13	H	N.A.
14	H	N.A.

[Next](#) [Up](#) [Previous](#)

**Up:** [Introduction to QM simulations](#) **Previous:** [Minimization and Structural Study of a Model System](#)

## Subsections

- [Introduction](#)
- [Proton Affinity P\(H<sub>2</sub>O\) of water](#)
- [Proton Affinity P\(CH<sub>3</sub>OH\) of Methanol](#)
- [Further studies](#)

# Determination of Proton Affinities

## Introduction

Proton transfer reactions are of great importance in chemistry and in biomolecular processes of living organisms. The latter include most enzymatically catalyzed reactions, e.g. ATP hydrolysis/synthesis by [F<sub>1</sub>F<sub>0</sub>-ATPsynthase](#). Furthermore, the protonation state of chemical groups, e.g. the side chains of amino acids, is fundamentally related to their biomolecular function. The deprotonation/protonation of a given chemical group can be represented schematically as  $A^- + H^+ \rightarrow AH$ . A measure of the probability of a chemical group to be protonated/unprotonated is given by the  $pK_a$  defined via  $pK_a = -\log [A^-][H^+]/[AH]$ . The  $pK_a$  of a protonable group strongly depends on its molecular environment. It is possible to measure the  $pK_a$  experimentally, but this is generally not an easy task.

In a gas phase environment there are two quantities which are used to describe the ability of a molecule to accept a proton. The first one is the gas phase basicity and is the negative of the free energy associated with the protonation/deprotonation reaction. A second and more frequently used quantity is the proton affinity  $P(A)$ , defined as the negative of the enthalpy change at standard condition (i.e. temperature and pressure). Computational *ab initio* approaches can provide reliable values for proton affinities, which is important since they are hard to determine experimentally. In the following you will determine the proton affinity of water and its dependence on the level of theory employed to perform the calculation (i.e. basis set, electron correlation). Using your experience with water, you will then be able to estimate the proton affinity of methanol, CH<sub>3</sub>OH.

The proton affinity for the reaction  $A^- + H^+ \rightarrow AH$  is defined as the negative of the reaction enthalpy at 298.15K, and hence (T: temperature, R: ideal gas constant)

$$P(A) = -\Delta H = -\Delta E + RT$$

The energy of a nonlinear polyatomic molecule can be approximated as

$$E(T) = E_{\text{rot}}(T) + E_{\text{trans}}(T) + \text{ZPE} + E'_{\text{vib}}(T) + E_{\text{ele}}$$

Here ZPE stands for the zero point energy of the normal modes. From statistical mechanics we know that the contributions  $E_{\text{rot}}$  and  $E_{\text{trans}}$  both equal  $3/2 RT$ . Furthermore,  $E_{\text{vib}}$  can usually be neglected compared to the zero point energy ZPE. To determine the proton affinity one has to calculate the energy change in going from the reactant  $A^-$  and  $H^+$  to the product  $AH$ . The rotational energy contribution remains constant (since the proton does not possess rotational kinetic energy) and the translational energy of the proton contributes  $-3/2 RT$ . Hence, neglecting the contribution due to the vibrations as argued above, one obtains the following expression for the proton affinity

$$P(A) = -\Delta E_{\text{ele}} - \Delta \text{ZPE} + 5/2 RT$$

To determine the proton affinity one therefore has to calculate two contributions: the change in electronic energy given by (note that the contribution of the proton is zero)

$$\Delta E_{\text{ele}} = E_{\text{ele}}(\text{H}_3\text{O}^+) - E_{\text{ele}}(\text{H}_2\text{O})$$

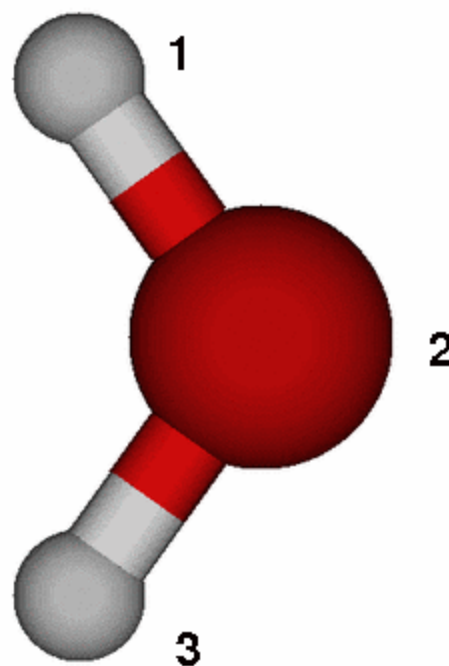
and the difference in zero point energies. In order to calculate the energies of  $AH$  and  $A^-$  you have to first optimize both systems. The ZPE can then be obtained by calculating the normal modes of the system via determination of the Hessian matrix.

## Proton Affinity $P(\text{H}_2\text{O})$ of water

You will start with a very simple, but important, test system and determine the proton affinity of water. During the first calculation you will employ Hartree-Fock theory and the 6-31G basis set. In the case of water one can schematically write  $\text{H}_2\text{O} + \text{H}^+ \rightarrow \text{H}_3\text{O}^+$ . The unprotonated/protonated species correspond to  $\text{H}_2\text{O}$  and  $\text{H}_3\text{O}^+$  respectively.

You therefore have to minimize both  $\text{H}_2\text{O}$  and  $\text{H}_3\text{O}^+$  and subsequently determine the electronic and the zero point energy at the minimum conformation.

First go into the working directory for part 2, and then into the directory for the proton affinity of water, i.e. `QM_tutorial/part2/pa_wat`. Please run the minimization of  $\text{H}_2\text{O}$  and  $\text{H}_3\text{O}^+$  in the directories `wat631` and `h3o631` respectively. After what you learned during the first part of the hands-on session you are now ready to set up the minimization runs yourself. You can use the input files for the minimization run of part 1 (the retinal analogue) as template and change the `$DATA` group as well as the number of internal coordinates in `NZVAR`. In order to construct the initial Z-matrix you may assume an O-H bond length of  $0.95 \text{ \AA}$ , a bond angle of 109 degrees and a dihedral of 120 degrees in the case of  $\text{H}_3\text{O}^+$ . It is useful if you make a schematic drawing of both molecules and use them to identify the internal coordinates. You can use the figure on the right as template. Please also add the keyword



HSEND=.TRUE. to the \$STATPT group to enforce evaluation of the Hessian matrix during the same computation, after the geometry search has converged. If you need help you can consider the sample files in the `sample_files` directory but we urge you to try it yourself and consider the GAMESS manual for help.

After you have completed setting up the two runs start your jobs using the `rungms` script. As a reminder how to go about doing that, type

```
tbss> rungms [inputfile] >& [outputfile]
```

They should both complete quickly and once they are finished make sure that they both terminated normally and the geometry optimization actually converged. Extract the final energies, which are output in the log files in the following form (you need the TOTAL ENERGY)

```
-----
ENERGY COMPONENTS
-----

WAVEFUNCTION NORMALIZATION =          1.0000000000

      ONE ELECTRON ENERGY =        -730.8096432790
      TWO ELECTRON ENERGY =         277.7469135169
      NUCLEAR REPULSION ENERGY =         204.9489137980
      -----
      TOTAL ENERGY =          -248.1138159640

ELECTRON-ELECTRON POTENTIAL ENERGY =         277.7469135169
NUCLEUS-ELECTRON POTENTIAL ENERGY =        -979.1237023677
NUCLEUS-NUCLEUS POTENTIAL ENERGY =         204.9489137980
      -----
      TOTAL POTENTIAL ENERGY =        -496.4278750527
      TOTAL KINETIC ENERGY =         248.3140590887
      VIRIAL RATIO (V/T) =           1.9991935893
```

Next, extract the zero point energies (look for HARMONIC ZERO POINT ENERGY and pick the kcal/mol value) which you can find at the bottom of the log file and which should look similar to

```
-----
THERMOCHEMISTRY AT T= 298.15 K
-----

USING IDEAL GAS, RIGID ROTOR, HARMONIC NORMAL MODE APPROXIMATIONS.
P= 1.01325E+05 PASCAL.
ALL FREQUENCIES ARE SCALED BY 1.00000
THE MOMENTS OF INERTIA ARE (IN AMU*BOHR**2)
      5.00304      5.00307      10.00611
THE ROTATIONAL SYMMETRY NUMBER IS 1.0
THE ROTATIONAL CONSTANTS ARE (IN GHZ)
      360.39877      360.39646      180.19881
THE HARMONIC ZERO POINT ENERGY IS (SCALED BY 1.000)
      0.036419 HARTREE/MOLECULE      7993.049924 CM**-1/MOLECULE
      22.853299 KCAL/MOL      95.618205 KJ/MOL
```

Using these values you can then calculate the proton affinity of water obtained via HF/6-31G (use  $T=298\text{K}$  and  $R=1.9872\text{ cal mol}^{-1}\text{ K}^{-1}$ ). First you need to determine  $\Delta E_{\text{ele}}$  and  $\Delta ZPE$

$$\Delta E_{\text{ele}} = E(\text{H}_3\text{O}^+) - E(\text{H}_2\text{O})$$

$$\Delta \text{ZPE} = \text{ZPE}(\text{H}_3\text{O}^+) - \text{ZPE}(\text{H}_2\text{O})$$

Note that since the energy difference  $\Delta E_{\text{ele}}$  is in Hartree (this is what GAMESS provides you with in the output file) you have to convert it to kcal/mol by multiplying it with a factor of 627.53 (1 Hartree = 627.53 kcal/mol). The difference in zero point energy  $\Delta \text{ZPE}$  already is in kcal/mol (since this is the value you picked from the log file). Finally, from the equation

$$P(A) = -\Delta E_{\text{ele}} - \Delta \text{ZPE} + 5/2RT$$

you can now calculate the proton affinity for water at an HF/6-31G level of theory.

Now that we have calculated the proton affinity using a rather "small" basis set, we need to find out the effect of a larger basis on the value of the proton affinity. In order to be able to estimate this you will next perform a similar calculation (i.e. use the same initial structure) employing a large basis set, namely the TZV triple zeta basis with additional polarization and diffuse functions. You can select this basis by changing your **\$BASIS** group to **\$BASIS GBASIS=TZV NDFUNC=1 NPFUNC=1 DIFFSP=.TRUE. DIFFS=.TRUE. \$END**. Using this basis set, perform minimizations and Hessian matrix calculations of  $\text{H}_2\text{O}$  and  $\text{H}_3\text{O}^+$ . Please use the directories `wattzv` and `h3otzv` for this purpose and consider the `sample_files` if you need help. Use this to obtain a second value for the proton affinity  $P(\text{H}_2\text{O})$  (i.e. calculate it!).

Don't forget to use MOLDEN to look at your system. Compare the final minimized structure with the one obtained using HF/6-31G, particularly the one of  $\text{H}_3\text{O}^+$ . You will observe that the HF/6-31 level of theory is not able to yield the correct non-planar structure of  $\text{H}_3\text{O}^+$ .

Next you will examine the effect of electron correlation on the calculated value of the proton affinity. You will use the HF/6-31G basis and the coupled cluster methodology with single and double excitations and triples-correction (CCSD(T)) to account for electronic correlation. Since the CCSD(T) method only allows single point energy calculations and no optimizations, you have to calculate the CCSD(T) energy at the final HF/6-31G minimized conformation. Please perform the calculations for  $\text{H}_2\text{O}$  and  $\text{H}_3\text{O}^+$  in the directories `wat631cc` and `h3o631cc`. For each  $\text{H}_2\text{O}$  and  $\text{H}_3\text{O}^+$  start from your 6-31G input files (this is the first calculation we did), replace the initial coordinates with the optimized ones (from 6-31G) and insert the keyword `CCTYP=CCSD(T)` in the `$CONTRL` group to request a CCSD(T) calculation. Don't forget to change `RUNTYPE` to `RUNTYPE=ENERGY` since we are only performing a single point run and remove `HSEND=.TRUE.` in the `$STATPT` group.

For simplicity you may assume that the ZPE for the CCSD(T) calculation is the same as for the initial HF/6-31G run, although this is, of course, an over-simplification. If you are not sure that your files are correct you may peek at the CCSD(T) input file in the `sample_file` directory. After your simulations have terminated successfully, extract the energies to calculate the proton affinity  $P(\text{H}_2\text{O})$  at the CCSD(T)//HF/6-31G level of theory.

Finally, you will calculate the proton affinity of water using the highest and most sophisticated level of theory available (at least in GAMESS), namely CCSD(T) with the TZV basis set. You should use the



TZV minimized structure as input for the CCSD(T) single point energy evaluation as described in the previous paragraph. This will provide the fourth and most refined value for  $P(\text{H}_2\text{O})$ . Please perform this calculations in the directories `h3o631cc` and `wattzvcc`.

Next, compare the 4 calculated values of  $P(\text{A})$  for the different levels of theory you considered above, as well as the individual contributions (i.e.  $\Delta E_{\text{ele}}$  and  $\Delta \text{ZPE}$ ) What is the effect of the size of the basis set and the correlation effects? The experimental value of  $P(\text{A})$  for water is  $P(\text{A})=165$  kcal/mol; which of your four calculations comes closest?

For the simple case of water, it was straightforward to calculate  $\Delta E_{\text{ele}}$  and  $\Delta \text{ZPE}$  even using a very large and computationally expensive level of theory like CCSD(T)//HF/TZV. Such a calculation will only be possible for the smallest systems and will become prohibitive even for medium sized molecules. A way to approximate the proton affinity for a larger molecule is to calculate the proton affinity at a lower level of theory (e.g. HF/6-31G) and then add corrections which approximately account for the basis set size and/or electron correlation. You should use the water example to calculate corrections C1 and C2 which arise due to a different basis set size and electron correlation effects respectively. To obtain correction C1 simply calculate the difference in  $P(\text{A})$  for the HF/6-31G and HF/TZV levels of theory; similarly subtract the  $P(\text{A})$  from HF/6-31G from CCSD(T)//HF/TZV to obtain C2. Hence:

$$\begin{aligned} \text{C1} &= P(\text{H}_2\text{O}, \text{HF}/\text{TZV}) - P(\text{H}_2\text{O}, \text{HF}/6-31\text{G}) \\ \text{C2} &= P(\text{H}_2\text{O}, \text{CCSD}(\text{T})//\text{HF}/6-31\text{G}) - P(\text{H}_2\text{O}, \text{HF}/6-31\text{G}) \end{aligned}$$

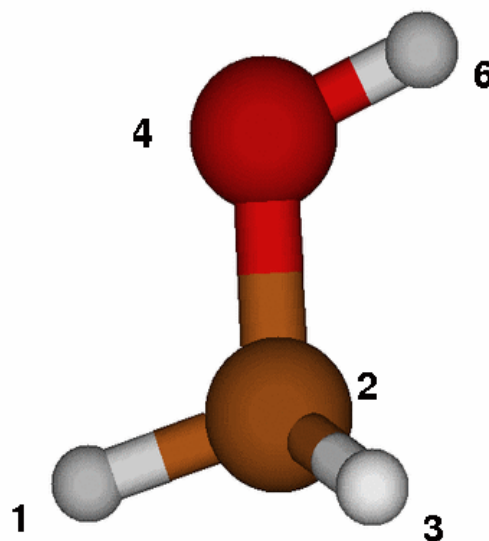
You can now estimate  $P(\text{H}_2\text{O}, \text{CCSD}(\text{T})//\text{HF}/\text{TZV})$  by adding corrections C1 and C2 to  $P(\text{H}_2\text{O}, \text{HF}/6-31\text{G})$ . How well does this result agree with the value you actually calculated in the last example? Why are they different and what assumptions have been made by simply assuming additivity? You can check you results for the proton affinities  $P(\text{A})$  and the correction factor C1 and C2 by clicking [here](#).

## Proton Affinity $P(\text{CH}_3\text{OH})$ of Methanol

In this section you will utilize the experience gained in calculating the proton affinity of water to estimate  $P(\text{CH}_3\text{OH})$ , the proton affinity of methanol.

All the calculations in this part will be performed in the directory `QM_tutorial/part2/pa_met` hence make sure to go there first. The water example has shown that you should use a high level of theory, e.g. CCSD(T)//HF/TZV to do proton affinity calculations. Since a simulation at this level of theory for methanol would take too long, you will have to determine  $P(\text{CH}_3\text{OH})$  at an HF/6-31G level of theory and then use the corrections C1 and C2 determined for water to take into account the larger basis set as well as electron correlation (this is what we did at the end of the last part for water).

The two species you have to consider for calculating  $\Delta E_{\text{ele}}$  and  $\Delta \text{ZPE}$  are the protonated and un-protonated



species, namely  $\text{CH}_3\text{OH}$  and  $\text{CH}_3\text{O}^+$ . Hence set up the input files to perform an energy minimization with subsequent Hessian matrix calculation at an HF/6-31G level of theory for both molecules. Please perform all calculations in the directories `ch3oh631` and `ch3o631`. To determine the Z-matrix you might want to sketch each molecule on a piece of paper and determine the bonds, angles and dihedrals you need to describe the system. You may assume that the  $\text{CH}_3$  group has a tetrahedral conformation with a C-H bond length of 1.1Å. The C-O bond length can be estimated at 1.4Å, the O-H bond with 0.95Å. Please try it yourself since mastering this task is very important. Use the figure on the left as template and employ the same atom numbering in the Z-matrix (atom number 5 is hidden behind the carbon!). In case you get stuck you may peek at the `sample_files`.

After you are done with the calculations you are ready to compute  $P(\text{CH}_3\text{OH})$  at the HF/6-31G level of theory and estimate the CCSD(T)//HF/TZV value by adding the correction C1 and C2 determined for water. We have computed the actual value of  $P(\text{CH}_3\text{OH})$  at the CCSD(T)//HF/TZV level for you and you should compare them to your estimated values. Do so by considering the pre-calculated files in `ch3ohcc` and `ch3occ` or simply by looking at the values given [here](#).

## Further studies

Of great interest, particularly in a biomolecular setting is the dependence of proton affinities on the molecular environment. So far you have assumed that the proton transfer takes place in the gas phase, i.e. in vacuum, which is of course not the case in a real system. To study the influence of the environment on the proton affinity you can, e.g., embed your molecule in a solvent environment or surround it by e.g. protein residues. The first task can be accomplished inside GAMESS using the `$PCM` input group which allows to immerse the system in a dielectric environment modeling the influence of solvent (c.f. the GAMESS manual). You can also design or pick from an appropriate pdb file part of a protein and study e.g. the proton affinity of water inside a cavity. The overall system size should still be small, otherwise the simulations will take a long time. There are many more issues to be examined and you are encouraged to explore them yourself.

---

[Next](#) [Up](#) [Previous](#)

**Up: Previous:**

*markus@ks.uiuc.edu*

Proton affinities of H2O for 4 levels of theory calculated via  
formula (\*) in the tutorial (all values in kcal/mol)

\*\*\*\*\*

level of theory	/\E(ele)	/\ZPE	P(A)
HF/6-31G	-182.60	8.74	175.34
HF/TZV	-176.71	8.61	169.58
CCSD(T)//HF/6-31G	-182.93	8.74	175.67
CCSD(T)//HF/TZV	-175.65	8.61	168.51

Computation of corrections C1 and C2 (in kcal/mol)

\*\*\*\*\*

$$C1 = 169.58 - 175.34 = -5.76$$

$$C2 = 175.67 - 175.34 = 0.32$$

Hence we can estimate P(H2O,CCSD(T)//HF/TZV) to be

$$P = P(\text{H2O, HF/6-31G}) + C1 + C2 = 175.34 - 5.76 + 0.32 = 169.90$$

which has to be compared with the "true" value of

$$P(\text{H2O, CCSD(T)//HF/TZV}) = 168.51$$

and is therefore in good agreement!

Proton affinities of methanol for 2 levels of theory calculated via formula (\*) in the tutorial (all values in kcal/mol)

\*\*\*\*\*

level of theory	/\E(ele)	/\ZPE	P(A)
HF/6-31G	-407.60	10.34	398.74
CCSD(T)//HF/TZV	-394.86	10.59	385.75

Estimate high level result P(CH3OH,CCSD(T)//HF/TZV) by applying corrections C1 and C2 to P(CH3OH,HF/6-31G) (in kcal/mol)

\*\*\*\*\*

$$\begin{aligned}
 P(\text{estimate}) &= P(\text{CH}_3\text{OH},\text{HF}/6-31\text{G}) + \text{C1} + \text{C2} \\
 &= 398.74 - 5.76 + 0.32 \\
 &= 393.30
 \end{aligned}$$

This has to be compared to the calculated value

$$P(\text{CH}_3\text{OH},\text{CCSD}(\text{T})//\text{HF}/\text{TZV}) = 385.75$$